

1 ANÁLISIS ESTADÍSTICO DE DATOS CATEGÓRICOS.

Ejemplo: Se realiza un estudio para conocer las preferencias sobre el tipo de gaseosa que se consume: "CC"=Coca Cola, "PC"=Pepsi Cola, "SC"=Salva Cola, para ello se toma una muestra aleatoria de 20 personas.

- 1) Activar el directorio de trabajo.
- 2) Crear un nuevo script y llamarle Script06-DatosCategoricos
- 3) Crear un vector con el tipo de gaseosa y otro con la muestra generada aleatoriamente:

```
tipo <- c("CC", "PC", "SC")
tipo

## [1] "CC" "PC" "SC"

# crea un vector en las que contiene los tres tipos de refrescos
consumo <- sample(tipo, 20, replace=TRUE)
consumo

## [1] "PC" "CC" "PC" "SC" "SC" "SC" "SC" "SC" "SC" "SC" "SC" "PC" "PC" "SC" "PC"
## [16] "PC" "PC" "CC" "PC" "PC"

# genera una muestra de tamaño 20 obtenida de los elementos del vector Tipo y
#los elementos se seleccionan con reemplazamiento

# Suponiendo que se quiere editar o agregar datos
data.entry(consumo)
```

- 4) Guarde el vector en un archivo de datos

```
# Guardar los datos en su directorio de trabajo
write(consumo, "Consumo.txt")
```

- 5) Eliminar los objetos que existen en el espacio de trabajo (Workspace)

```
ls()

## [1] "consumo" "tipo"

rm(list=ls(all=TRUE))
ls()

## character(0)
```

- 6) Leer o recuperar el vector de datos o archivo de texto

```
Consumo <- scan("Consumo.txt", what = character(), na.strings = "NA",
               flush=FALSE)
Consumo

## [1] "PC" "CC" "PC" "SC" "SC" "SC" "SC" "SC" "SC" "SC" "SC" "PC" "PC" "SC" "PC"
## [16] "PC" "PC" "CC" "PC" "PC"

ls()

## [1] "Consumo"

# Si el vector contiene caracteres se ocupa: what = character() na.strings = "NA",
#le indica a R que los valores faltantes son identificados con \NA"
```

- 7) Crear la tabla de distribución de frecuencias y proporciones

```
frec <- table(Consumo)
frec

## Consumo
## CC PC SC
## 2 9 9

prop <- table(Consumo)/length(Consumo)
prop

## Consumo
## CC PC SC
## 0.10 0.45 0.45

# Note que la salida por defecto no es para nada atractiva en comparación con
#el resto de paquetes estadísticos
```

En cambio, si estamos usando LATEX y queremos incorporar estos cuadros o cualquier otro podemos utilizar el comando `xtable(table(Consumo))` (NOTE QUE EL ARGUMENTO DEBE SER UN CUADRO), y con esto automáticamente se nos genera el código en LATEX y luego incorporarlo a nuestro informe, lo mejor de todo es que salida resultante es mucho más presentable.

8) Conocer un resumen de los datos

```
summary(Consumo)

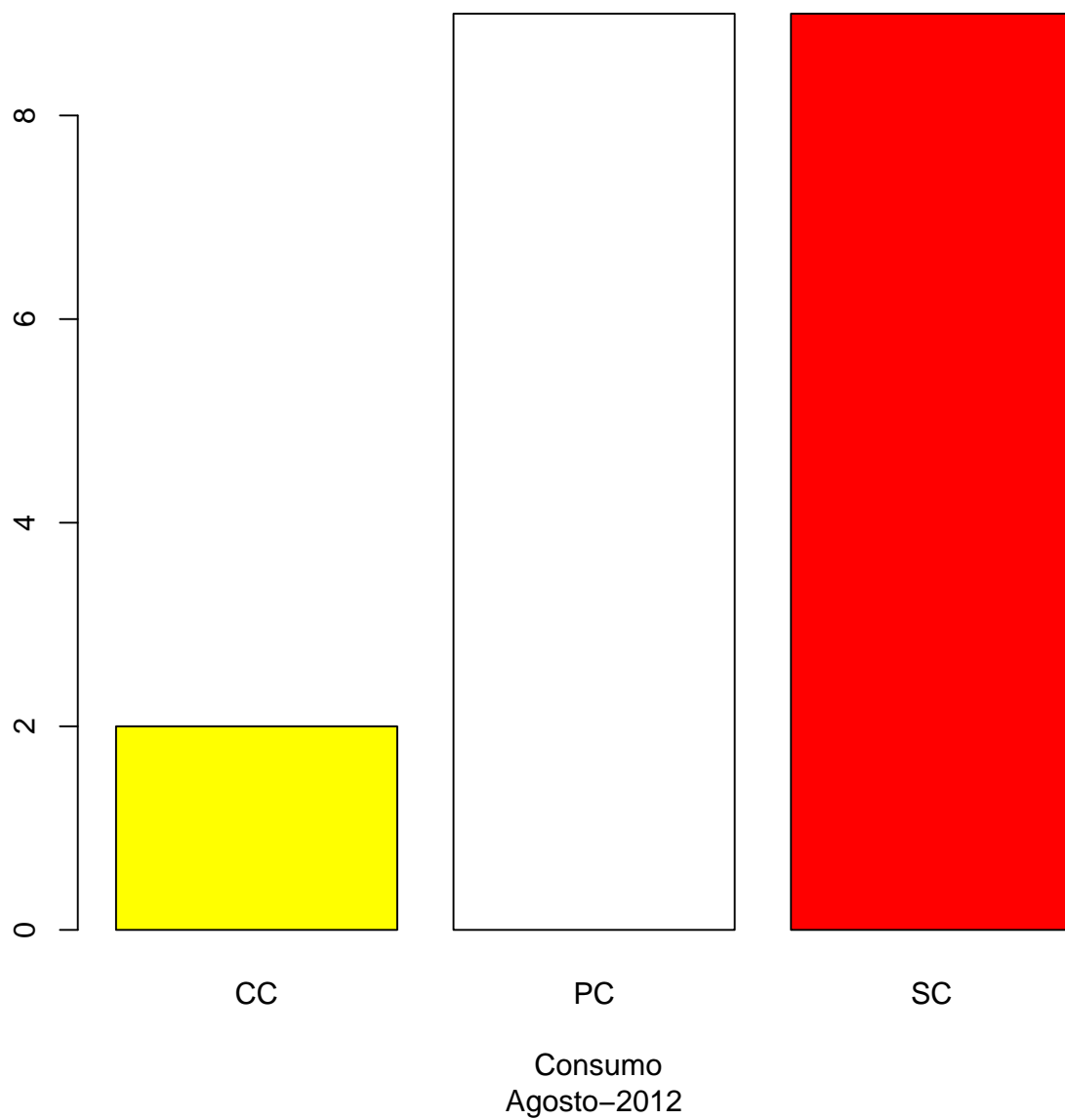
##      Length      Class      Mode
##           20 character character

# note que por tratarse de variables cualitativas únicamente muestra el número
#de elementos, y el tipo de datos.
```

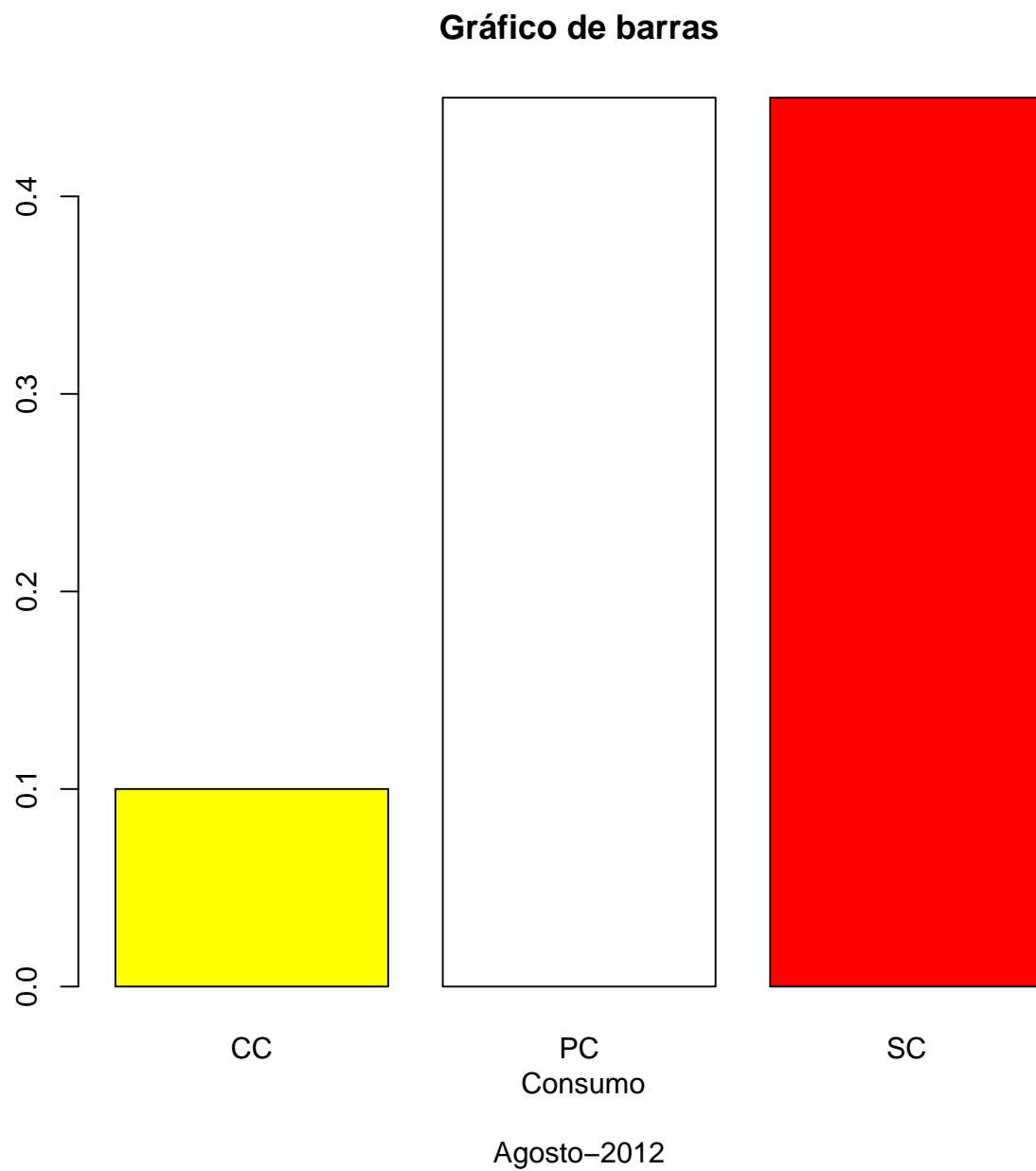
9) Realizar un gráfico de barras

```
# Para las frecuencias absolutas
barplot(frec, main="Gráfico de barras", xlab=" Consumo",
        col=c("yellow", "white", "red"), sub="Agosto-2012")
```

Gráfico de barras



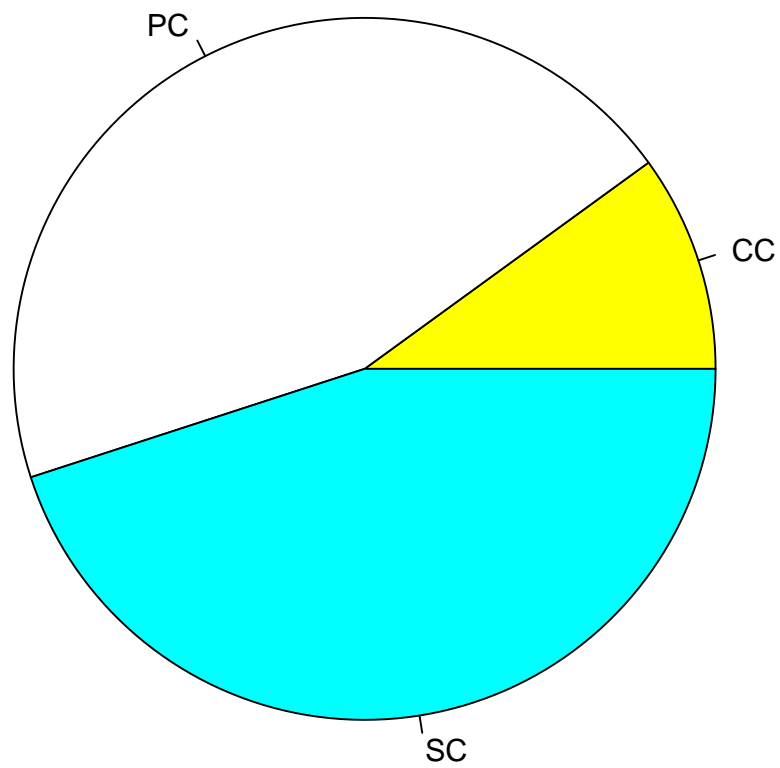
```
# Para las frecuencias relativas
barplot(prop, main="Gráfico de barras", xlab=" Consumo\n",
        col=c("yellow", "white", "red"), sub="Agosto-2012")
```



10) Realizar un gráfico de pastel

```
pie(frec, main="Gráfico de pastel", xlab="Tipo de Consumo",  
    col=c("yellow", "white", "cyan"), sub="Agosto-2012")
```

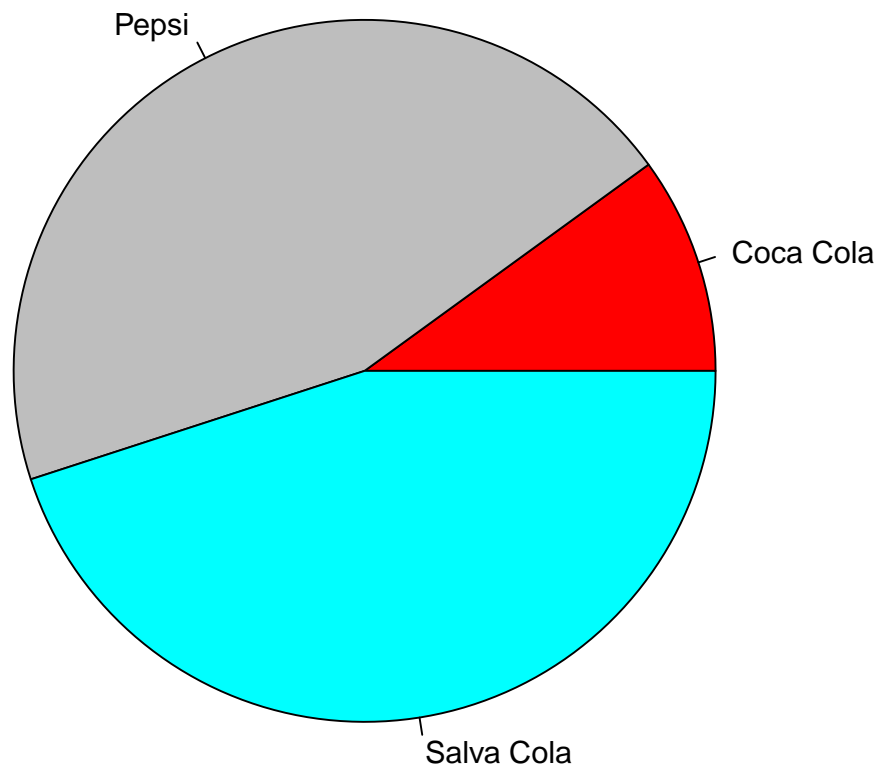
Gráfico de pastel



Tipo de Consumo
Agosto-2012

```
# Se puede especificar nombres para las categorías y el color de los sectores
names(frec) = c("Coca Cola", "Pepsi", "Salva Cola")
pie(frec, main="Gráfico de pastel", xlab=" Consumo", radius=0.8,
    col=c("red", "gray", "cyan"), sub="Agosto-2012")
```

Gráfico de pastel



Consumo
Agosto-2012

*#Los colores se asignan dependiendo del orden en que han sido especificados por names()
#Note con la instrucción radius se especifica el tamaño de la figura, mientras
#más cerca de uno (uno de menos uno) se encuentre más grande será (el ángulo cambia).*

11) Colocar valores numéricos en los sectores del gráfico

```
n <- length(frec)
hoj <- data.frame(frec)
hoj

##      Var1 Freq
## 1 Coca Cola  2
## 2 Pepsi     9
## 3 Salva Cola 9

etiq <- c(paste(hoj$Var1, "-", hoj$Freq)); etiq

## [1] "Coca Cola - 2" "Pepsi - 9"      "Salva Cola - 9"

pie(frec, main="Gráfico de pastel", labels=etiq, col=rainbow(n), border=TRUE)
```

Gráfico de pastel

